Hepatoprotective effects of systemic ER activation

BulkRNAseq - Transcriptome molecular signatures

Christian Sommerauer & Carlos Gallardo

16 August, 2022

```
# source and library import
source('code/00_helper_functions.R')
library(tidyverse)
library(DESeq2)
```

```
# color palettes
colPals <- list()
colPals$conditions <- setNames(c('#E98BB6', '#B02262', '#7F9AD7', '#2A2F72', '#7DC7D1', '#339ACD', '#35
                               c('CDf', 'HFDf', 'CDm', 'HFDm', 'DPN', 'DIP', 'E2', 'PPT'))
colPals$RdBu <- rev(RColorBrewer::brewer.pal(n=11, name = 'RdBu'))
colPals$UpDown <- setNames(colPals$RdBu[c(10,2)],
                           c('up', 'down'))
```

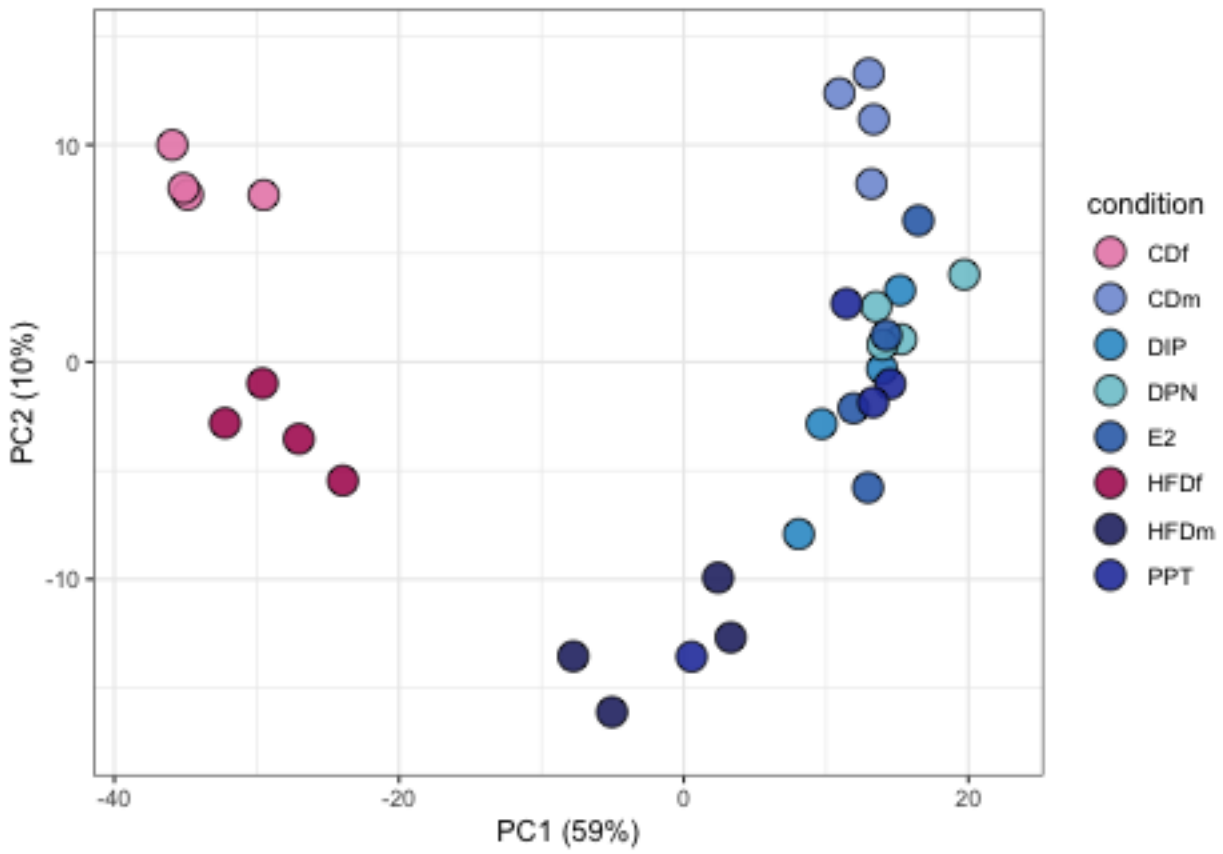
Load data

```
# consensus differentially expressed genes
DEGs <- readRDS('results/bulkRNAseq_mmus_DEGs.rds')

# raw counts RNAseq
raw_counts <- read.table(
  file = 'data/bulkRNAseq_mmus_rawcounts.tsv',
  stringsAsFactors = FALSE,
  sep = '\t',
  header = TRUE) %>%
  dplyr::filter(geneID %in% DEGs$unfilt$CDfVsCDm$ensembl_gene_id) %>%
  tibble::column_to_rownames('geneID') %>%
  as.matrix()

# gene lengths
gene_len <- read.table(
  file = 'data/bulkRNAseq_mmus_gene_lengths.tsv',
  stringsAsFactors = FALSE,
  sep = '\t',
  header = TRUE) %>%
  dplyr::filter(geneID %in% DEGs$unfilt$CDfVsCDm$ensembl_gene_id)

# design RNAseq
```

```r
design_meta <- read.table(
  file = 'data/bulkRNAseq_mmus_design.tsv',
  stringsAsFactors = FALSE,
  sep = '\t',
  header = TRUE)

# ensembl gene annotation (Mus musculus)
gene_ann <- read.table(
  file = 'data/ensembl_mmus_sep2019_annotation.tsv',
  stringsAsFactors = FALSE,
  sep = '\t',
  header = TRUE,
  fill = FALSE,
  quote = '') %>%
  dplyr::filter(ensembl_gene_id %in% DEGs$unfilt$CDfVsCDm$ensembl_gene_id) %>%
  dplyr::arrange(factor(ensembl_gene_id, levels = rownames(raw_counts)))
```

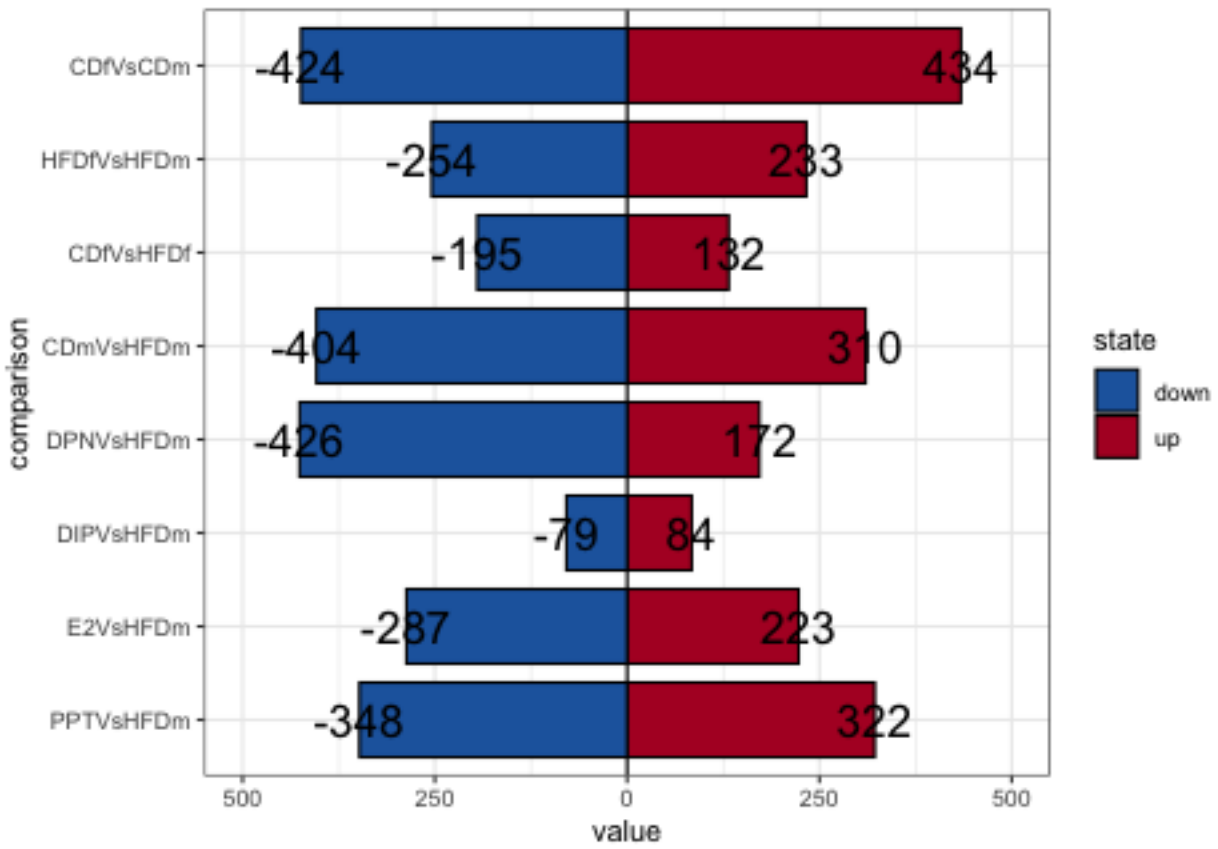
Principal component analysis (PCA)

```r
pca_res <- DESeq2::DESeqDataSetFromMatrix(countData = raw_counts,
                                  colData = design_meta,
                                  design = ~0 + condition) %>%
  DESeq2::estimateSizeFactors() %>%
  DESeq2::DESeq() %>%
  DESeq2::vst(blind = FALSE) %>%
  assay() %>%
  doPCA()

df <- data.frame(PC1 = pca_res$pcs$PC1,
                 PC2 = pca_res$pcs$PC2,
                 condition = design_meta$condition)

ggplot(df, aes(x=PC1, y=PC2, fill=condition),) +
  geom_point(shape=21, size=5, stroke=0.5, color='black') +
  scale_fill_manual(values = alpha(colPals$conditions, 0.9)) +
  scale_x_continuous(expand = expansion(mult = c(.1, .1))) +
  scale_y_continuous(
    expand =expansion(mult = c(.1, .1))) +
  xlab(paste0('PC1 (', round(pca_res$perc_var[1]), '%)')) +
  ylab(paste0('PC2 (', round(pca_res$perc_var[2]), '%)')) +
  theme_bw()
```



Differentially expressed genes (DEGs)

```
up <- lapply(DEGs$filt, function(x) sum(x$log2FoldChange>0)) %>% unlist()
down <- lapply(DEGs$filt, function(x) sum(x$log2FoldChange<0)) %>% unlist()

df <- data.frame(comparison=factor(rep(names(DEGs$filt),2), levels=names(DEGs$filt)),
                 state=c(rep('up', length(up)), rep('down', length(down))),
                 value=c(up, down*-1))

ggplot(df, aes(x=comparison, y=value, fill=state, label=value)) +
  geom_hline(yintercept = 0, linetype='solid', size=0.5) +
  geom_bar(color='black', size=0.5, width=0.8, position='stack', stat='identity') +
  geom_text(size=6) +
  scale_fill_manual(values = colPals$UpDown) +
  scale_x_discrete(limits = rev) +
  scale_y_continuous(limits = c(-500,500), labels = c(500,250,0,250,500)) +
  coord_flip() +
  theme_bw()
```

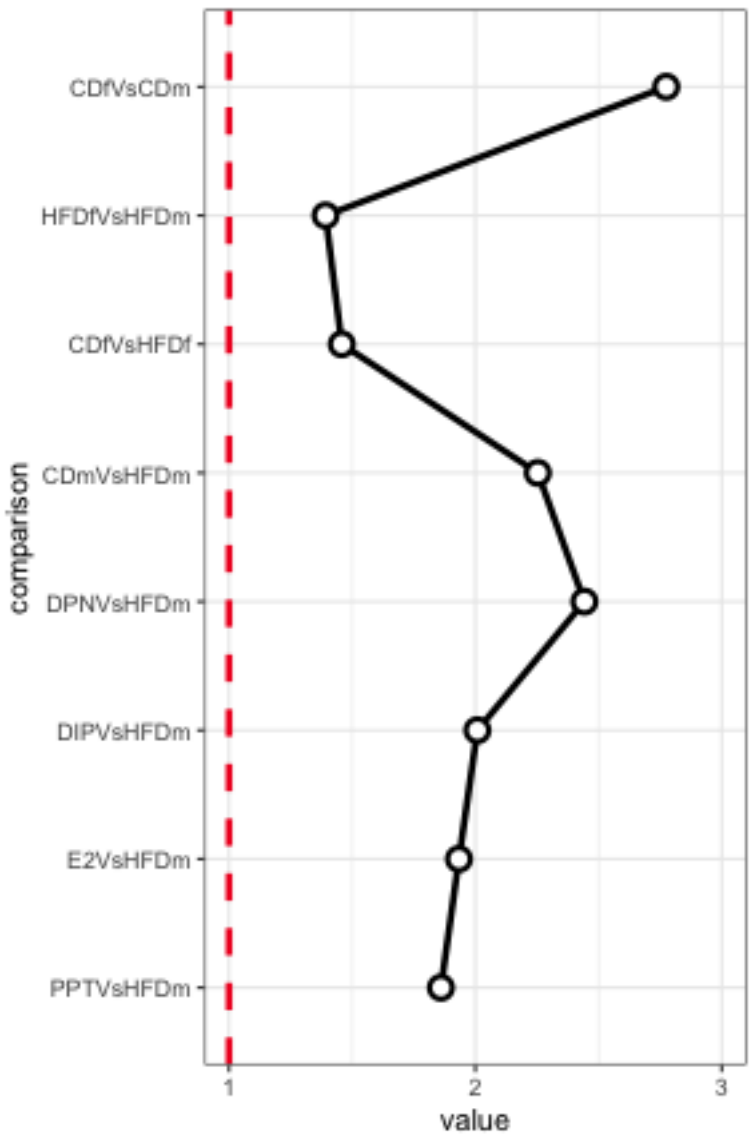


Filter and normalize RNAseq data

```
RNAseq <- list()

# remove outlier sample PPT_HFD_male_4
RNAseq$counts <- raw_counts %>%
  as.data.frame() %>%
  dplyr::select(-PPT_HFD_male_4)

RNAseq$design_meta <- design_meta %>%
  dplyr::filter(sample != 'PPT_HFD_male_4')

RNAseq$annotation <- gene_ann %>%
  dplyr::rename(geneID = ensembl_gene_id) %>%
  dplyr::left_join(gene_len, by = 'geneID')

RNAseq$cpm <- RNAseq$counts %>%
  normalizeData(method = 'CPM')

RNAseq$tpm <- RNAseq$counts %>%
  normalizeData(len = RNAseq$annotation$length, method = 'TPM')
```

Transcriptome-wide signal-to-noise ratios (tSNR)

```
df <- RNAseq$tpm %>%
  scaleData(method = 'zscore') %>%
  tSNR(group.lbls = RNAseq$design_meta$condition) %>%
  tibble::rownames_to_column(var = 'X') %>%
  tidyr::pivot_longer(cols = dplyr::everything()[-1], names_to = 'Y') %>%
  tidyr::unite(col = 'comparison', X, Y, sep = 'Vs') %>%
  dplyr::filter(comparison %in% names(DEGs$filt)) %>%
  dplyr::mutate(comparison=factor(comparison, levels = names(DEGs$filt)))

ggplot(df, aes(x=comparison, y=value)) +
  geom_line(group=1, size=1.2) +
  geom_point(shape=21, size=3, stroke=1.5, color='black', fill='white') +
  geom_hline(yintercept = 1, linetype='dashed', size=1.2, color='#EF2126') +
  scale_x_discrete(limits = rev) +
  scale_y_continuous(limits = c(1,3), breaks = c(1,2,3)) +
  coord_flip() +
  theme_bw()
```

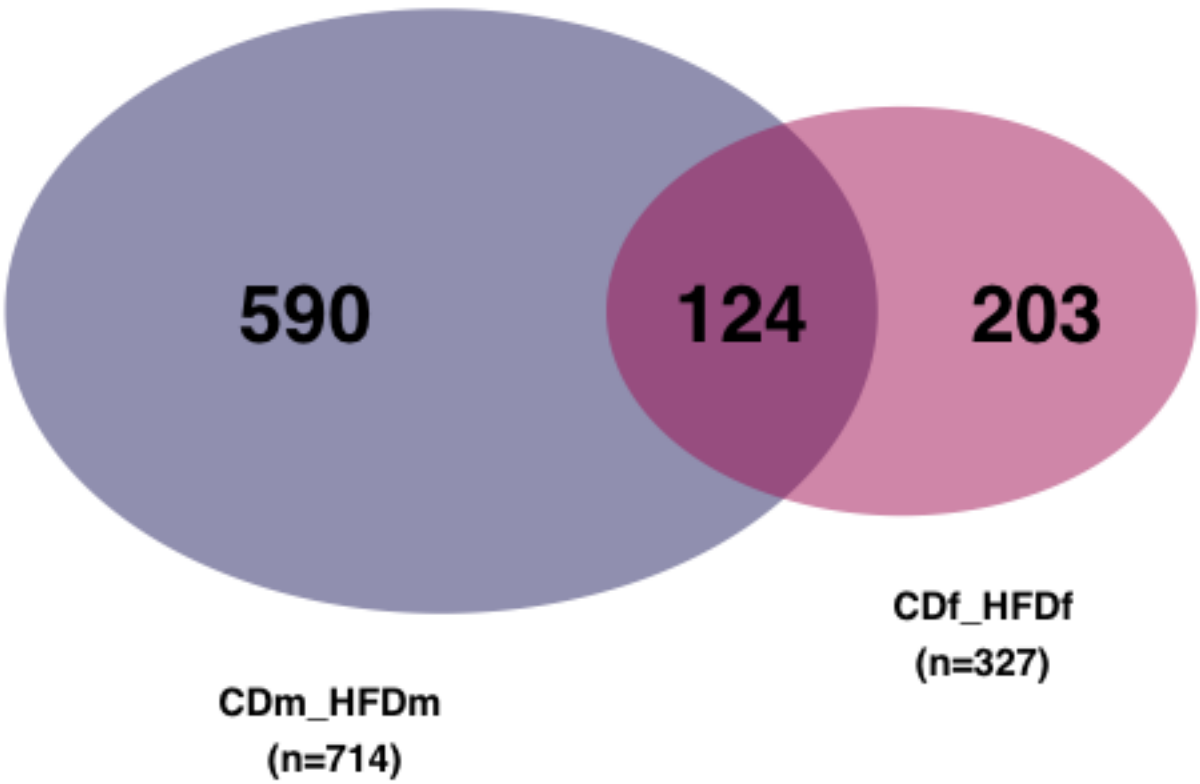


TBA Fig. S1E-J

Plot Venn Diagram (Fig S1E)

```
library(tidyverse)
library(VennDiagram)

CDf_HFDf <- DEGs$filt$CDfVsHFDf
CDm_HFDm <- DEGs$filt$CDmVsHFDm

grid.newpage()
myCol <- c("#B02262", "#2A2E72")
venn.plot <- venn.diagram(
```

```
  x = list(CDf_HFDf$ensembl_gene_id, CDm_HFDm$ensembl_gene_id),
  category.names = c("CDf_HFDf\n(n=327)", "CDm_HFDm \n(n=714)"),
  filename = NULL,
  output=T,
  lwd = 2,
  lty = 'blank',
  fill = myCol,
    cex = 2.5,
        fontface = "bold",
        fontfamily = "sans",
  cat.cex = 1.2,
        cat.fontface = "bold",
        cat.default.pos = "outer",
        cat.fontfamily = "sans",
  cat.pos = c(-10, 10),
        cat.dist = c(0.08, 0.08))

grid.draw(venn.plot)
```

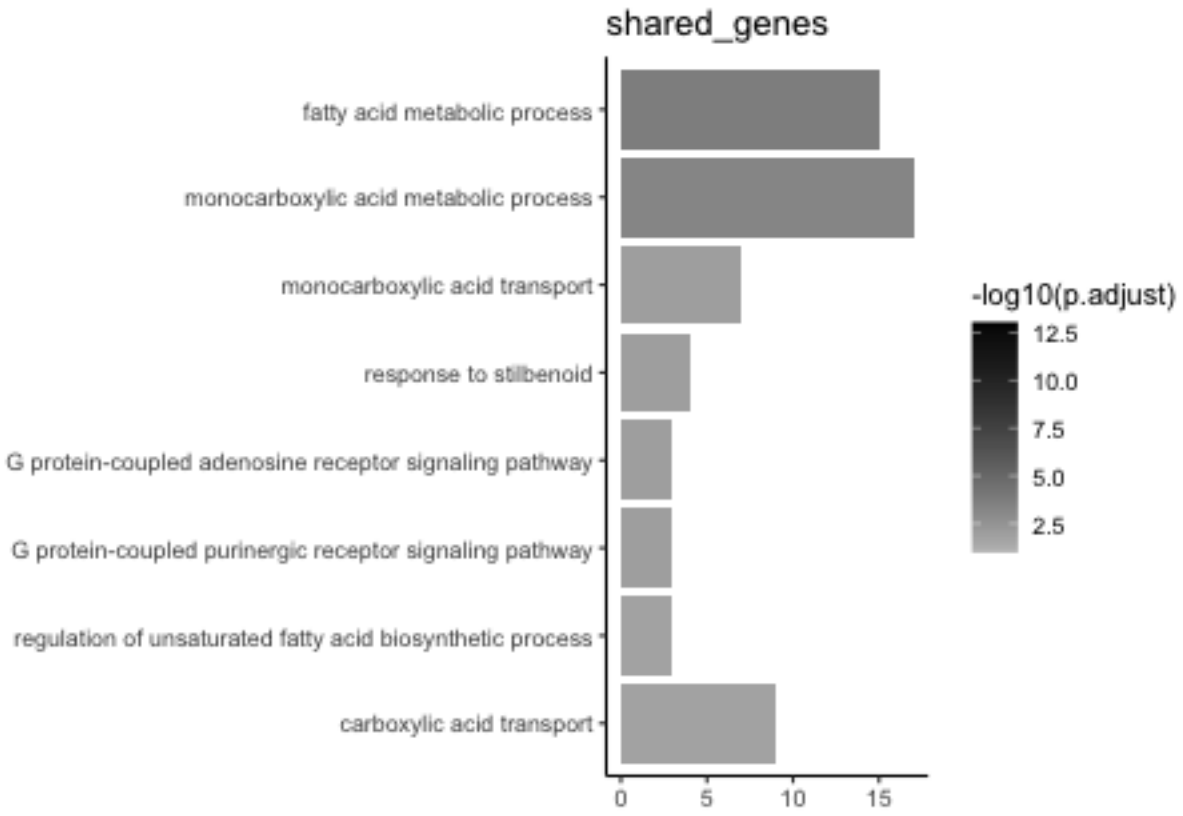


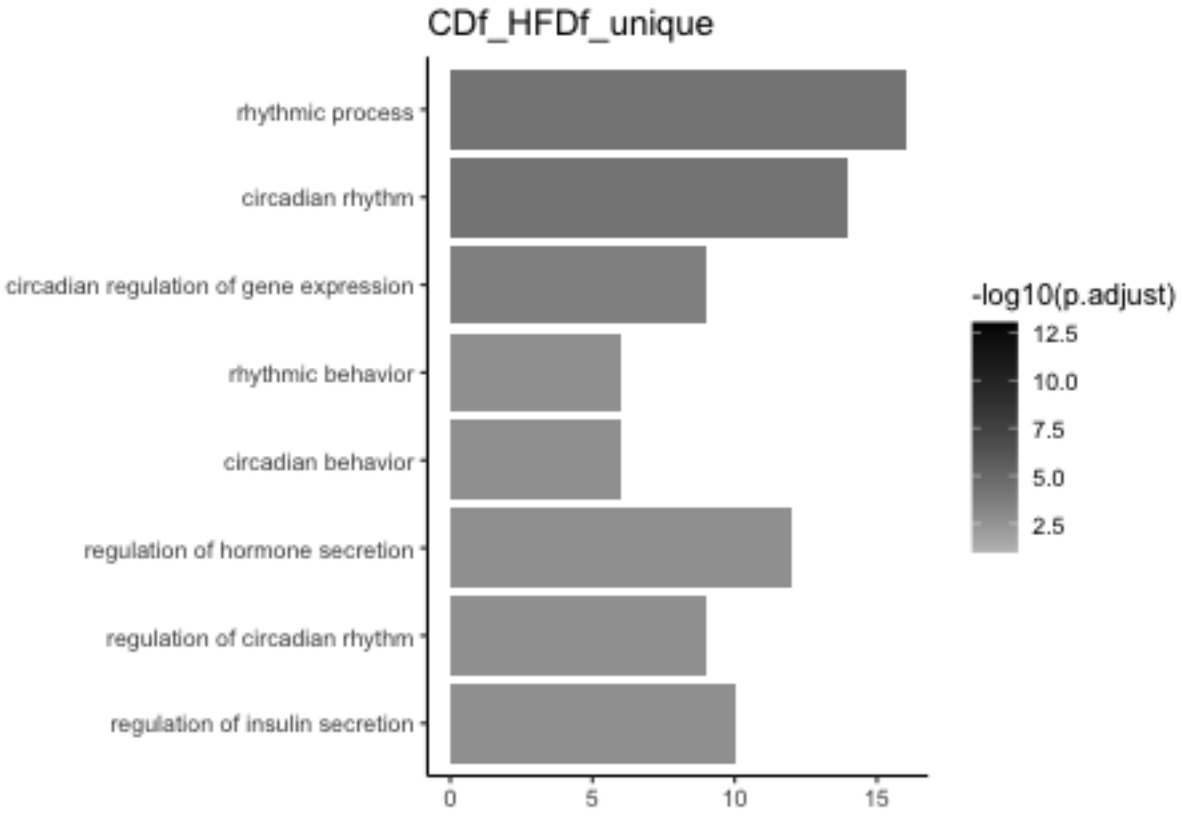

```
# Make intersection to extract the genes names from Venn.
intersections <- list("", "", "")
names(intersections) <- c("shared_genes", "CDf_HFDf_unique", "CDm_HFDm_unique")

intersections[[1]] <- intersect(CDf_HFDf$ensembl_gene_id,CDm_HFDm$ensembl_gene_id)
intersections[[2]] <- setdiff(CDf_HFDf$ensembl_gene_id, intersections$shared_genes)
```

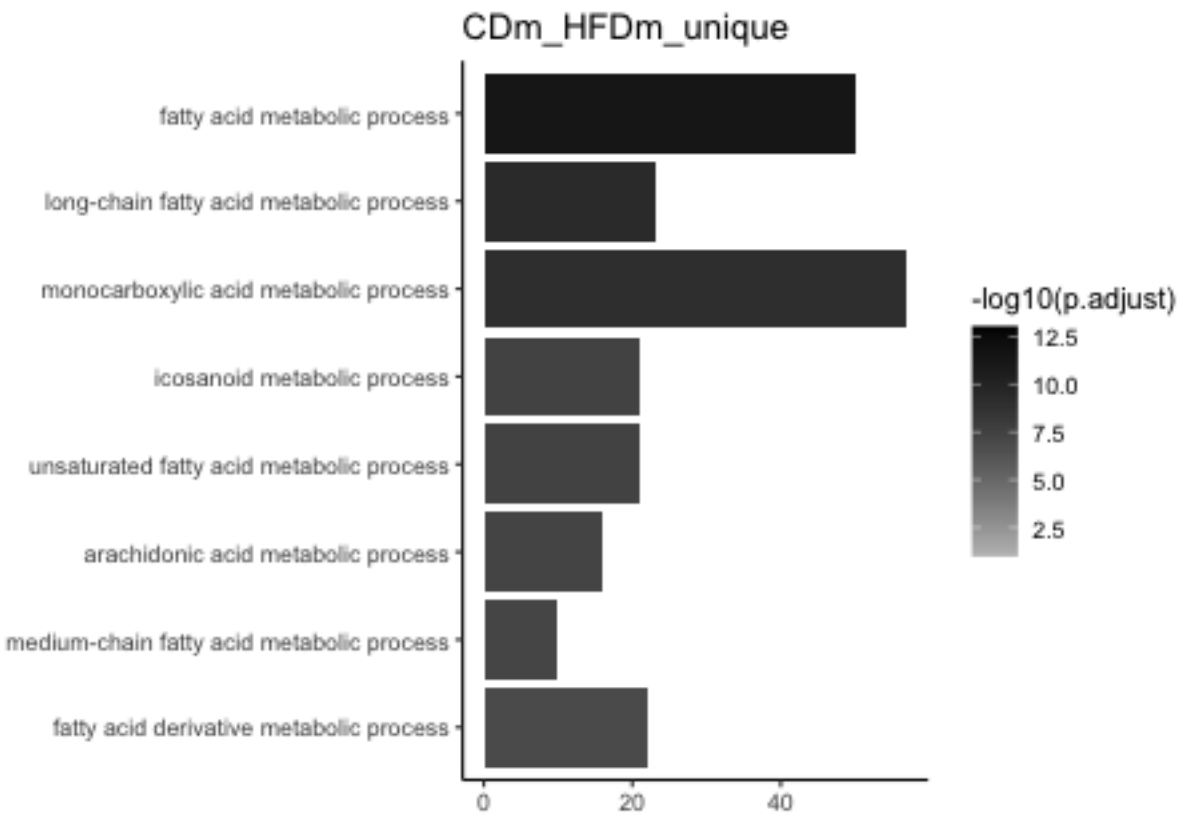
```
intersections[[3]] <- setdiff(CDm_HFDm$ensembl_gene_id, intersections$shared_genes)
```

Gene Ontologies (Fig S1F,G,H)

```
library(clusterProfiler)
background <- DEGs$unfilt$CDfVsCDm$ensembl_gene_id
options(connectionObserver = NULL) # workaround due to a bug
library(org.Mm.eg.db)

GO_list <- list(GO.results = list(),
                GO.top8 = list(),
                term.order.plotting = list())

for (i in 1:3) {
GO_list$GO.results[[i]] <- enrichGO(gene = intersections[[i]],
                keyType       = 'ENSEMBL',
                OrgDb         = org.Mm.eg.db,
                ont           = "BP",
                pAdjustMethod = "BH",
                pvalueCutoff  = 0.05,
                qvalueCutoff  = 0.05,
                minGSSize     = 3,
                readable      = TRUE,
                universe = background)
head(GO_list$GO.results[[i]])

name.me <- c("shared_genes", "CDf_HFDf_unique", "CDm_HFDm_unique")

GO_list$GO.top8[[i]] <- GO_list$GO.results[[i]]@result %>% filter(p.adjust<0.05) %>% mutate(GeneSet = na

GO_list$term.order.plotting[[i]] <- GO_list$GO.top8[[i]] %>% dplyr::pull("Description")

print(ggplot(GO_list$GO.top8[[i]], aes(x=Count, y=factor(Description, levels=rev(GO_list$term.order.plo
  geom_col(aes(fill=-log10(p.adjust))) +
  theme_classic() +
  xlab("") +
  ylab("") +
  ggtitle(paste(unique(GO_list$GO.top8[[i]]$GeneSet))) +
  scale_fill_gradient(low="grey", high= "black", limits=c(1,13)))

}
```


shared_genes
fatty acid metabolic process
monocarboxylic acid metabolic process
monocarboxylic acid transport
response to stilbenoid
G protein-coupled adenosine receptor signaling pathway
G protein-coupled purinergic receptor signaling pathway
regulation of unsaturated fatty acid biosynthetic process
carboxylic acid transport
0
5
10
15
-log10(p.adjust)
12.5
10.0
7.5
5.0
2.5


CDf_HFDf_unique
rhythmic process
circadian rhythm
circadian regulation of gene expression
rhythmic behavior
circadian behavior
regulation of hormone secretion
regulation of circadian rhythm
regulation of insulin secretion
0
5
10
15
-log10(p.adjust)
12.5
10.0
7.5
5.0
2.5



Plot volcanos (Fig S1I,J)

```
library(tidyverse)
library(ggrepel)

# Load dataframes from rds object.
## DEG in CD comparison
CDf_CDm <- DEGs$filt$CDfVsCDm

## All genes in given comparisons with fold-changes
CDf_CDm_unfilt <- DEGs$unfilt$CDfVsCDm
HFDf_HFDm_unfilt <- DEGs$unfilt$HFDfVsHFDm

# CD comparison
## Maximum significance and fold-change are capped here.
CDf_CDm$padj[CDf_CDm$padj < 10e-30] <- 10e-30
CDf_CDm$log2FoldChange[CDf_CDm$log2FoldChange < (-8)] <- (-8)
CDf_CDm$log2FoldChange[CDf_CDm$log2FoldChange > (8)] <- (8)

CDf_CDm_unfilt$padj[CDf_CDm_unfilt$padj < 10e-30] <- 10e-30
CDf_CDm_unfilt$log2FoldChange[CDf_CDm_unfilt$log2FoldChange < (-8)] <- (-8)
CDf_CDm_unfilt$log2FoldChange[CDf_CDm_unfilt$log2FoldChange  > (8)] <- (8)
```

```r
# HFD comparison
## We do not need the HFD DEG, because we will plot the CDf vs CDm DEG with the HFD fold-changes

HFDf_HFDm_unfilt$padj[HFDf_HFDm_unfilt$padj < 10e-30] <- 10e-30
HFDf_HFDm_unfilt$log2FoldChange[HFDf_HFDm_unfilt$log2FoldChange < (-8)] <- (-8)
HFDf_HFDm_unfilt$log2FoldChange[HFDf_HFDm_unfilt$log2FoldChange  > (8)] <- (8)

# Create up and downregulated genes in respective comparisons (larger/smaller than 0 is sufficient beca
male_biased_CD <- filter(CDf_CDm, log2FoldChange < 0)
nrow(male_biased_CD)
```

```
## [1] 424
```

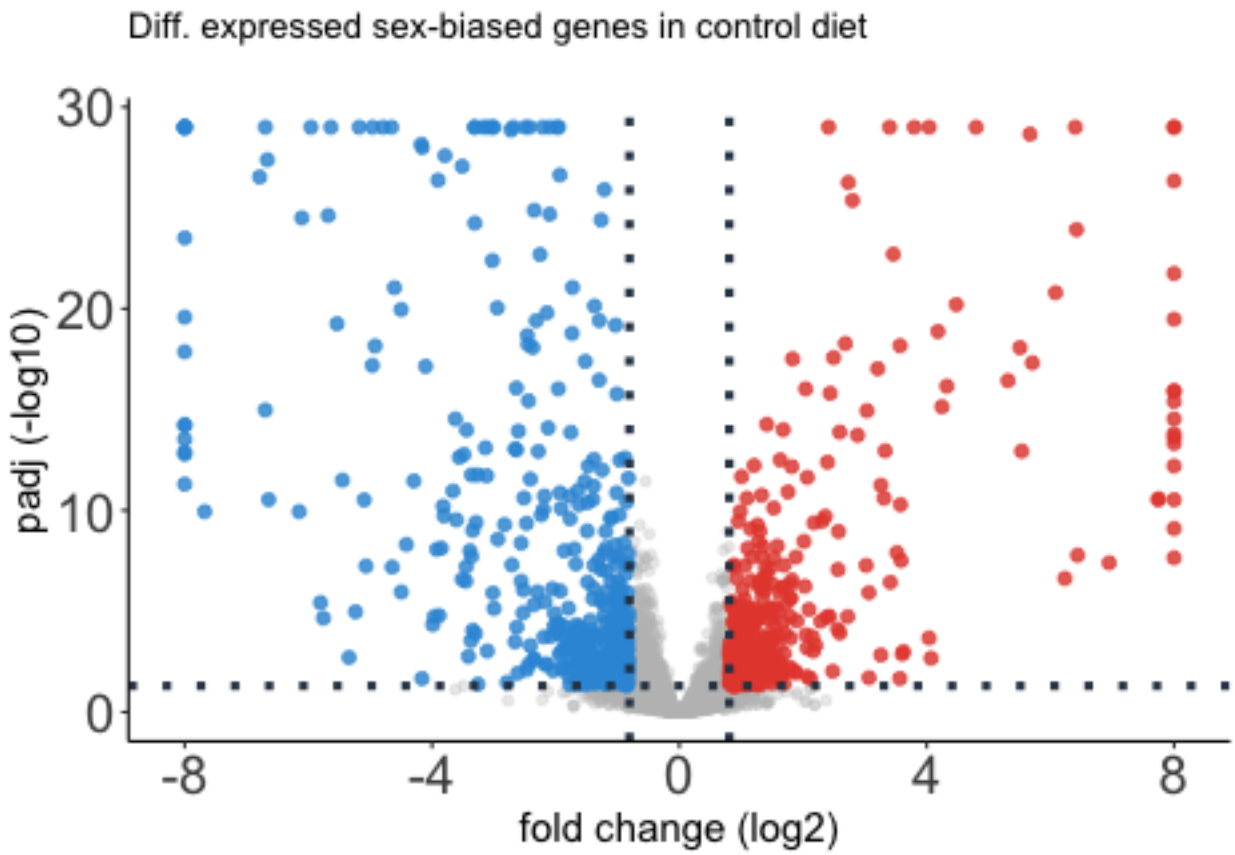
```r
female_biased_CD <- filter(CDf_CDm, log2FoldChange > 0)
nrow(female_biased_CD)
```

```
## [1] 434
```

```r
# These genes will be plotted in both, HFD and CD volcano plots, as sex-biased gene set.
male_biased_CD_vector <- filter(CDf_CDm, log2FoldChange < 0) %>% dplyr::pull("external_gene_name")
female_biased_CD_vector <- filter(CDf_CDm, log2FoldChange > 0) %>% dplyr::pull("external_gene_name")
# Filter the HFD table for genes that are sex-biased in CD.
male_biased_CD_in_HFDbckgrd <- HFDf_HFDm_unfilt %>% filter(external_gene_name%in%male_biased_CD_vector)
female_biased_CD_in_HFDbckgrd <- HFDf_HFDm_unfilt %>% filter(external_gene_name%in%female_biased_CD_vec
```

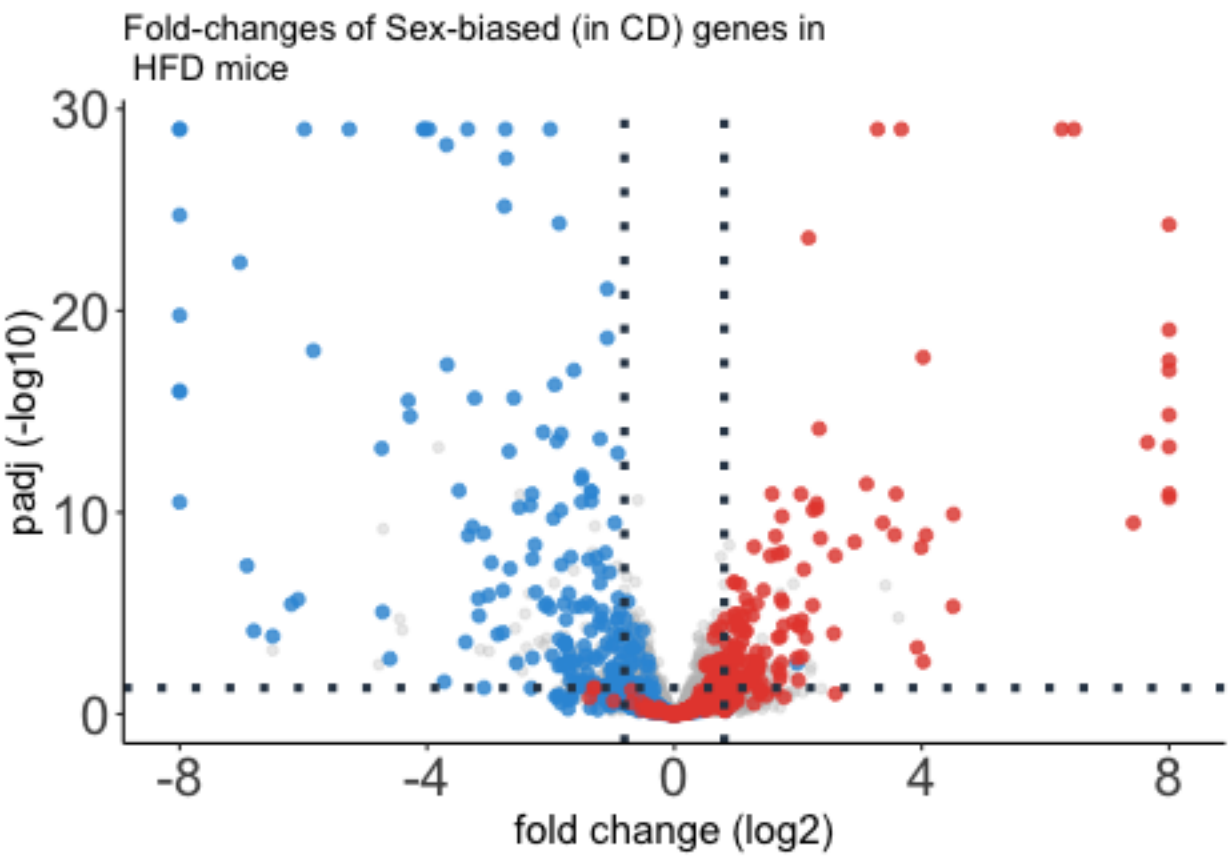
```r
ggplot(CDf_CDm_unfilt) +
  geom_point(data = CDf_CDm_unfilt,
    aes(x = log2FoldChange, y = -log10(padj)),
    color = "grey",
    alpha = 0.3,
    cex = 1.5) +
  geom_point(data = male_biased_CD,
    aes(x = log2FoldChange, y = -log10(padj)),
    color = "#3498db", # #3498db is blue
    alpha = 0.8,
    cex = 2) +
  geom_point(data = female_biased_CD,
    aes(x = log2FoldChange, y = -log10(padj)),
    color =  "#e74c3c" ,
    alpha = 0.8,
    cex = 2) +
  theme_classic() +
  theme(axis.text = element_text(size=20),
        axis.title.x = element_text(size=15),
        axis.title.y = element_text(size=15)) +
  scale_x_continuous(limits = c(-8.1, 8.1), breaks =c(-8, -4, 0, 4, 8)) +
  xlab("fold change (log2)") +
  ylab("padj (-log10)") +
  geom_vline(xintercept = 0.807,
    col = "#2e4053",
    linetype = "dotted",
    size = 1.5) +
```

```
  geom_vline(xintercept = -0.807,
    col = "#2e4053",
    linetype = "dotted",
    size = 1.5) +
  geom_hline(yintercept = -log10(0.05),
    col = "#2e4053",
    linetype = "dotted",
    size = 1.5) +
  ggtitle("Diff. expressed sex-biased genes in control diet\n")
```




```
ggplot(HFDf_HFDm_unfilt) +
  geom_point(data = HFDf_HFDm_unfilt,
    aes(x = log2FoldChange, y = -log10(padj)),
    color = "grey",
    alpha = 0.3,
    cex = 1.5) +
  geom_point(data = male_biased_CD_in_HFDbckgrd,
    aes(x = log2FoldChange, y = -log10(padj)),
    color = "#3498db",
    alpha = 0.8,
    cex = 2) +
  geom_point(data = female_biased_CD_in_HFDbckgrd,
    aes(x = log2FoldChange, y = -log10(padj)),
    color =  "#e74c3c" ,
    alpha = 0.8,
```

```
  cex = 2) +
theme_classic() +
theme(axis.text = element_text(size=20),
      axis.title.x = element_text(size=15),
      axis.title.y = element_text(size=15)) +
xlab("fold change (log2)") +
ylab("padj (-log10)") +
scale_x_continuous(limits = c(-8.1, 8.1), breaks =c(-8, -4, 0, 4, 8)) +
geom_vline(xintercept = 0.807,
  col = "#2e4053",
  linetype = "dotted",
  size = 1.5) +
geom_vline(xintercept = -0.807,
  col = "#2e4053",
  linetype = "dotted",
  size = 1.5) +
geom_hline(yintercept = -log10(0.05),
  col = "#2e4053",
  linetype = "dotted",
  size = 1.5) +
ggtitle("Fold-changes of Sex-biased (in CD) genes in \n HFD mice")
```



Export filtered and normalized RNAseq data

```
saveRDS(RNAseq, file = 'results/bulkRNAseq_mmus_data_filt_norm.rds')
```

```
sessionInfo()
```

```
## R version 4.0.3 (2020-10-10)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Big Sur 10.16
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRblas.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
##  [1] grid      parallel  stats4    stats     graphics  grDevices utils
##  [8] datasets  methods   base
##
## other attached packages:
##  [1] ggrepel_0.9.1               org.Mm.eg.db_3.12.0
##  [3] AnnotationDbi_1.52.0        clusterProfiler_3.18.1
##  [5] VennDiagram_1.6.20          futile.logger_1.4.3
##  [7] DESeq2_1.30.1               SummarizedExperiment_1.20.0
##  [9] Biobase_2.50.0              MatrixGenerics_1.2.1
## [11] matrixStats_0.58.0          GenomicRanges_1.42.0
## [13] GenomeInfoDb_1.26.7         IRanges_2.24.1
## [15] S4Vectors_0.28.1            BiocGenerics_0.36.1
## [17] forcats_0.5.1               stringr_1.4.0
## [19] dplyr_1.0.6                 purrr_0.3.4
## [21] readr_1.4.0                 tidyr_1.1.3
## [23] tibble_3.1.2                ggplot2_3.3.3
## [25] tidyverse_1.3.1
##
## loaded via a namespace (and not attached):
##   [1] fgsea_1.16.0           colorspace_2.0-1       ellipsis_0.3.2
##   [4] qvalue_2.22.0          XVector_0.30.0         fs_1.5.0
##   [7] rstudioapi_0.13        farver_2.1.0           graphlayouts_0.7.1
##  [10] bit64_4.0.5            scatterpie_0.1.5       fansi_0.5.0
##  [13] lubridate_1.7.10       xml2_1.3.2             splines_4.0.3
##  [16] cachem_1.0.5           GOSemSim_2.16.1        geneplotter_1.68.0
##  [19] knitr_1.33             polyclip_1.10-0        jsonlite_1.7.2
##  [22] broom_0.7.6            annotate_1.68.0        GO.db_3.12.1
##  [25] dbplyr_2.1.1           ggforce_0.3.3          BiocManager_1.30.12
##  [28] compiler_4.0.3         httr_1.4.2             rvcheck_0.1.8
##  [31] backports_1.2.1        assertthat_0.2.1       Matrix_1.3-3
##  [34] fastmap_1.1.0          cli_3.2.0              tweenr_1.0.2
##  [37] formatR_1.9            htmltools_0.5.1.1      tools_4.0.3
##  [40] igraph_1.2.6           gtable_0.3.0           glue_1.6.2
##  [43] GenomeInfoDbData_1.2.4 reshape2_1.4.4         DO.db_2.9
```

```
##  [46] enrichplot_1.10.2      fastmatch_1.1-0        Rcpp_1.0.6
##  [49] cellranger_1.1.0       vctrs_0.3.8            ggraph_2.0.5
##  [52] xfun_0.31              rvest_1.0.0            lifecycle_1.0.0
##  [55] XML_3.99-0.6           DOSE_3.16.0            MASS_7.3-53.1
##  [58] zlibbioc_1.36.0        scales_1.1.1           tidygraph_1.2.0
##  [61] hms_1.1.0              lambda.r_1.2.4         RColorBrewer_1.1-2
##  [64] yaml_2.2.1             gridExtra_2.3          memoise_2.0.0
##  [67] downloader_0.4         stringi_1.6.2          RSQLite_2.2.6
##  [70] highr_0.9              genefilter_1.72.1      BiocParallel_1.24.1
##  [73] rlang_0.4.11           pkgconfig_2.0.3        bitops_1.0-7
##  [76] evaluate_0.14          lattice_0.20-41        labeling_0.4.2
##  [79] shadowtext_0.0.7       cowplot_1.1.1          bit_4.0.4
##  [82] tidyselect_1.1.1       plyr_1.8.6             magrittr_2.0.1
##  [85] R6_2.5.0               generics_0.1.0         DelayedArray_0.16.3
##  [88] DBI_1.1.1              pillar_1.6.1           haven_2.4.1
##  [91] withr_2.4.2            survival_3.2-10        RCurl_1.98-1.3
##  [94] modelr_0.1.8           crayon_1.4.1           futile.options_1.0.1
##  [97] utf8_1.2.1             rmarkdown_2.14         viridis_0.6.1
## [100] locfit_1.5-9.4         readxl_1.3.1           data.table_1.14.0
## [103] blob_1.2.1             reprex_2.0.0           digest_0.6.27
## [106] xtable_1.8-4           munsell_0.5.0          viridisLite_0.4.0
```